Classification aided two stage localization

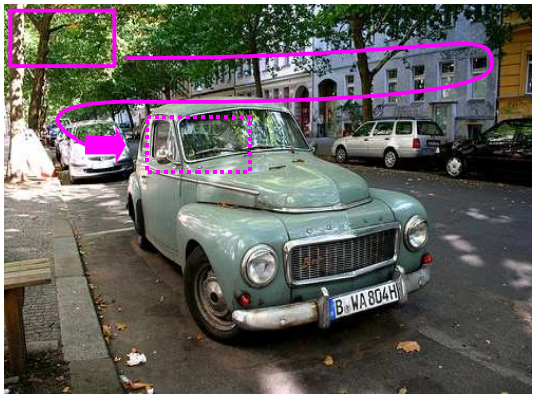
Hedi Harzallah, Cordelia Schmid, Frederic Jurie, Adrien Gaidon

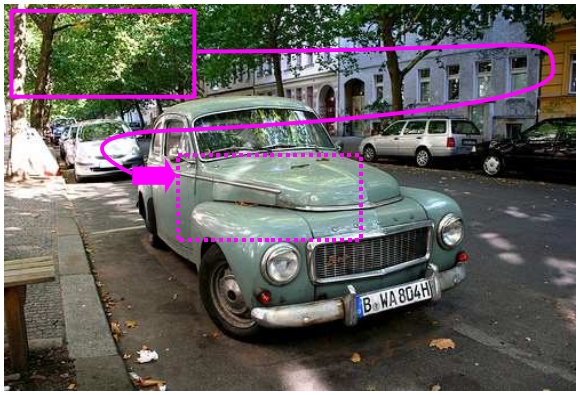
INRIA, LEAR project team

PASCAL VOC2008 challenge workshop

Introduction

- Detection task on PASCAL VOC2008 challenge
- Method with sliding windows (Each window is classified as containing or not the targeted object)



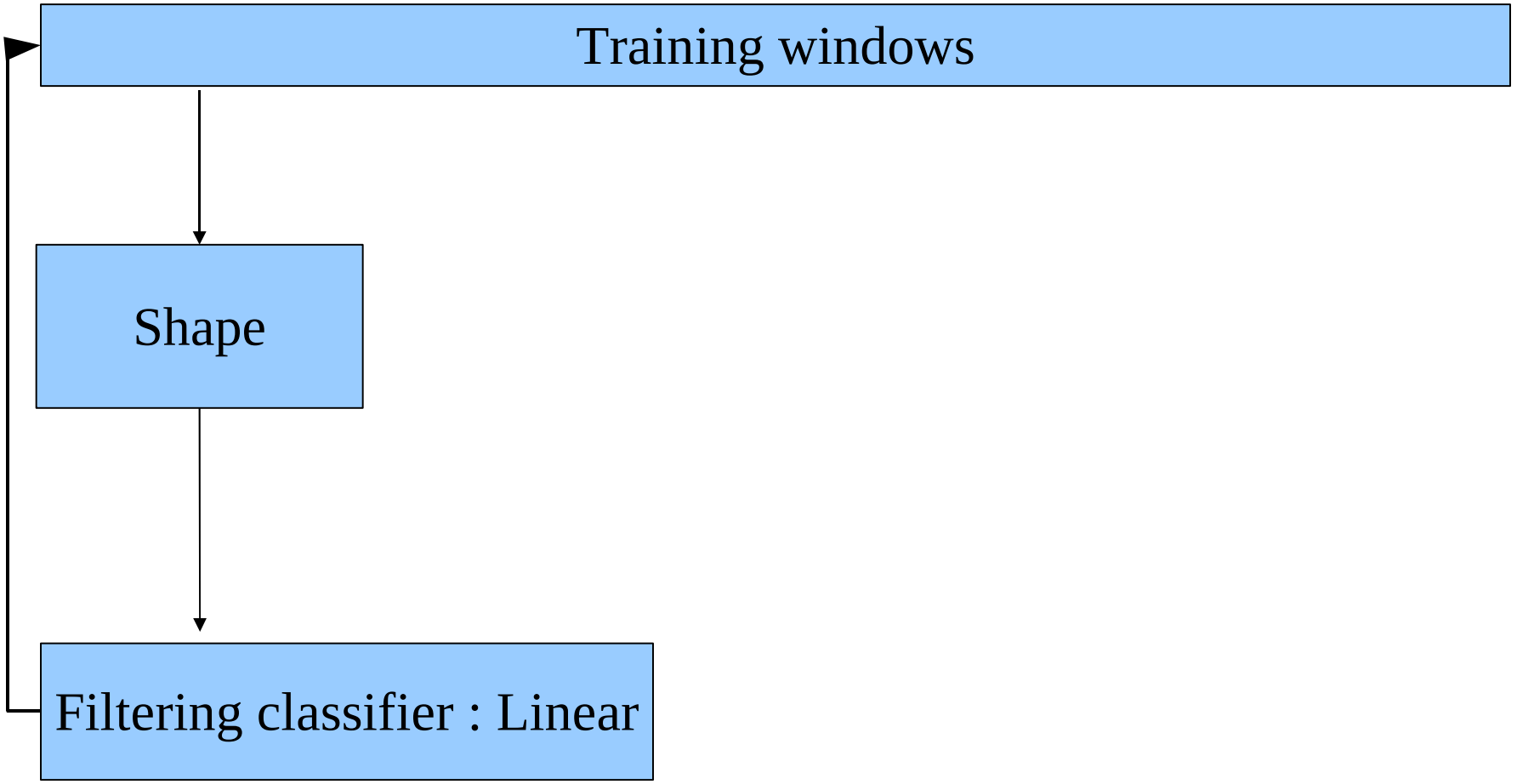


- Learn a classifier by providing positive and negative examples

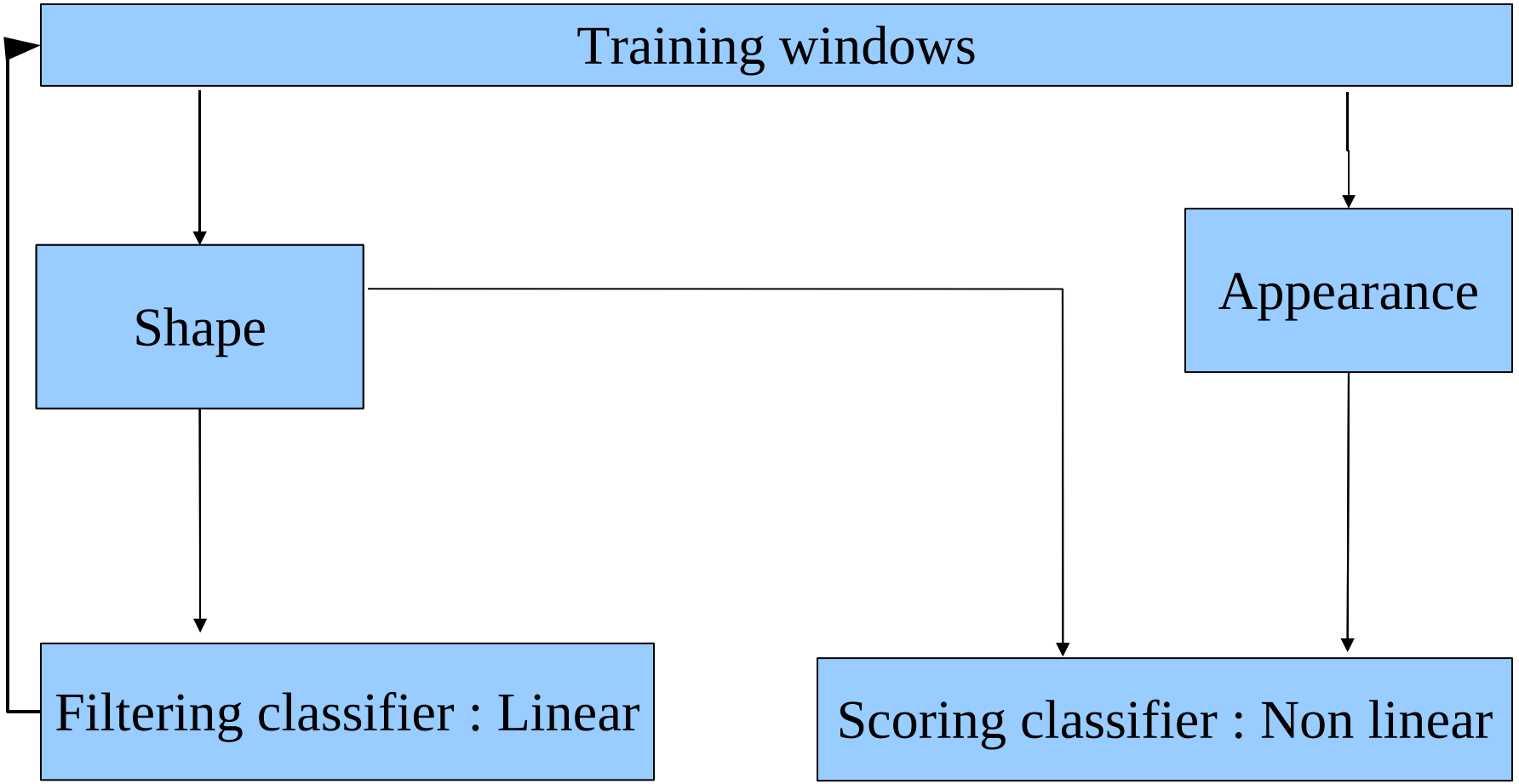
INRIA

LEAR

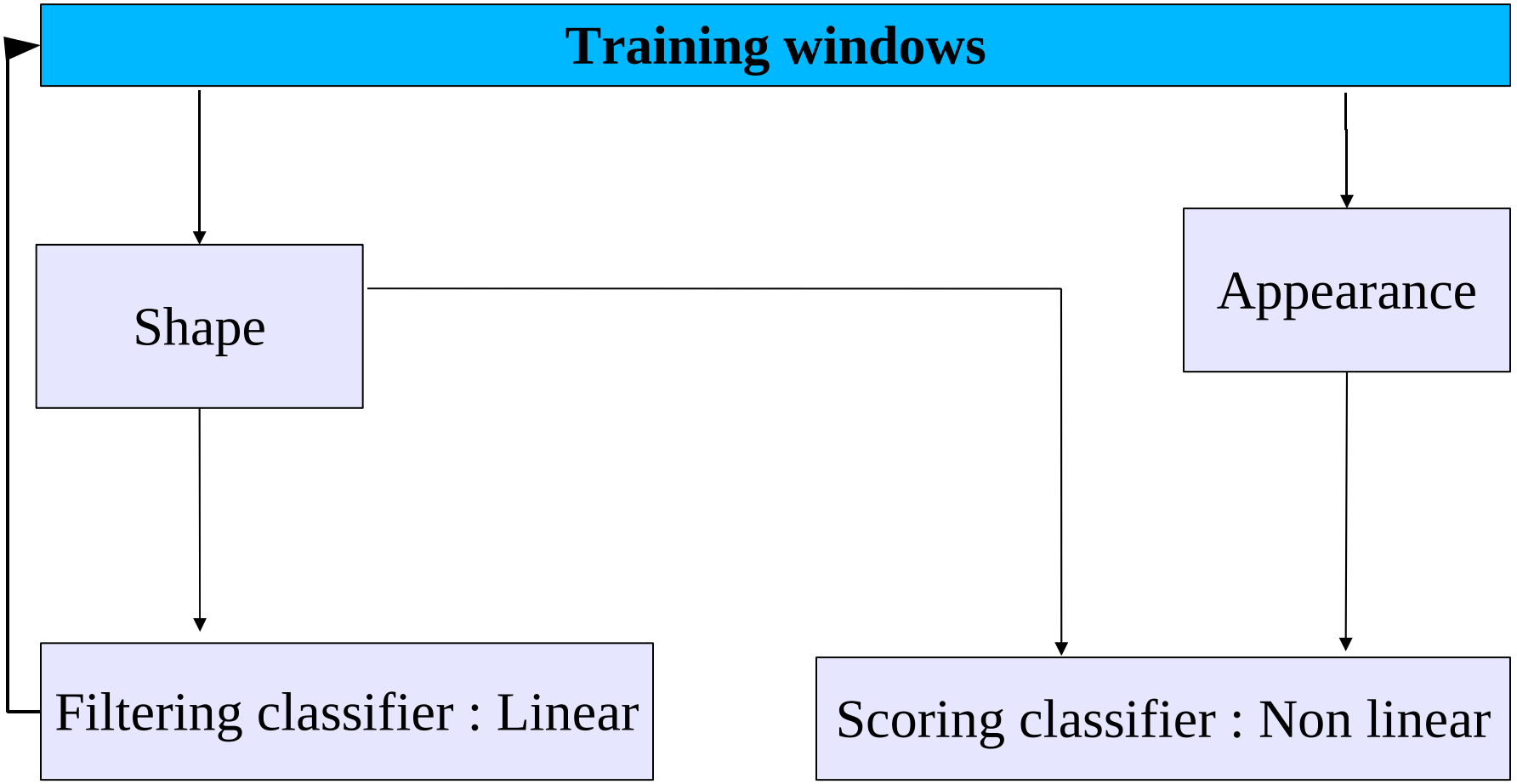
Training outline



Training outline

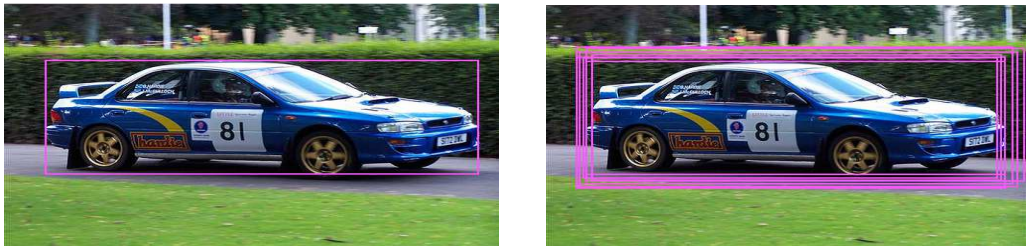


Generating training windows



Generating training windows

- Adding positive training examples by shifting and scaling the original annotations [Laptev06]



- Negative examples randomly extracted from background
- Training an initial classifier
- Retraining 4 times by adding false positives

Examples of false positives

Image representation

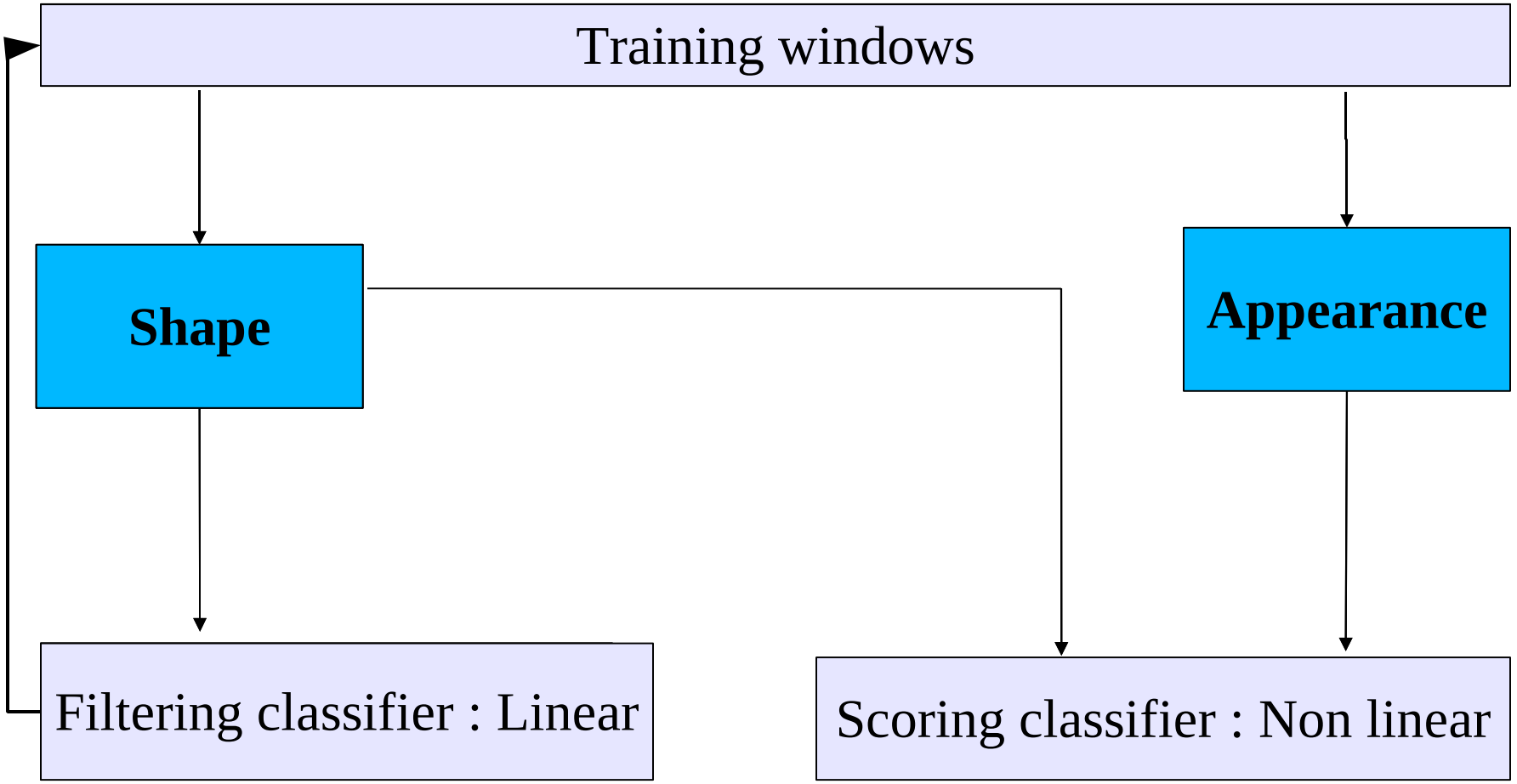
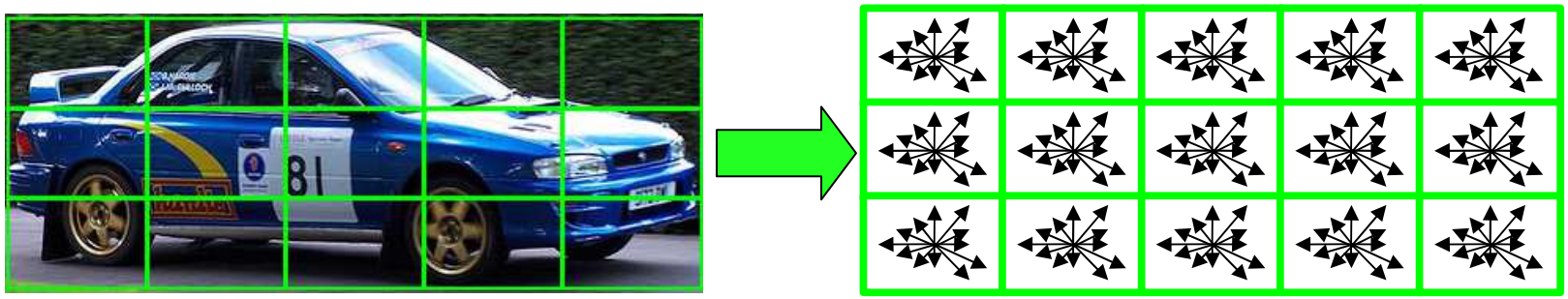


Image representation: Histogram-of-gradients (HOG)

- Tiling optimized per class (around 100 overlapping tiles)



- Computed with integral histograms

- With 16 orientations

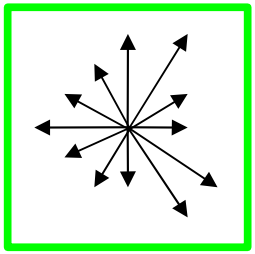
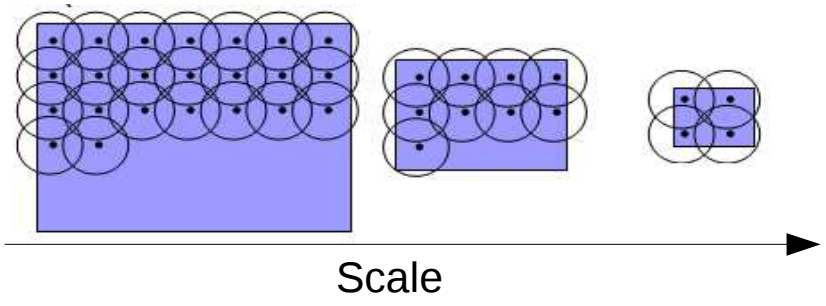
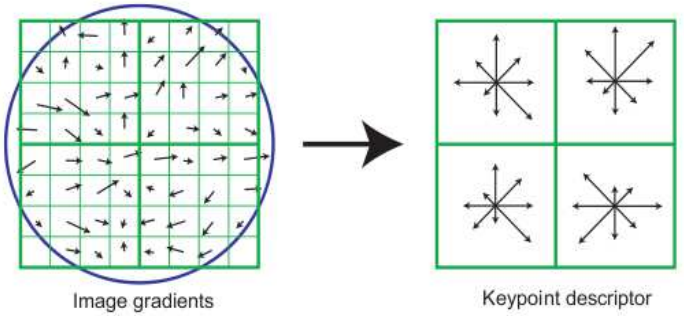
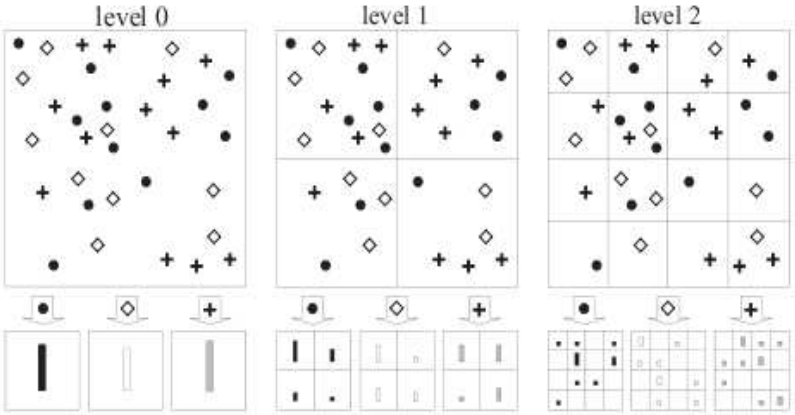


Image representation : Dense SIFT





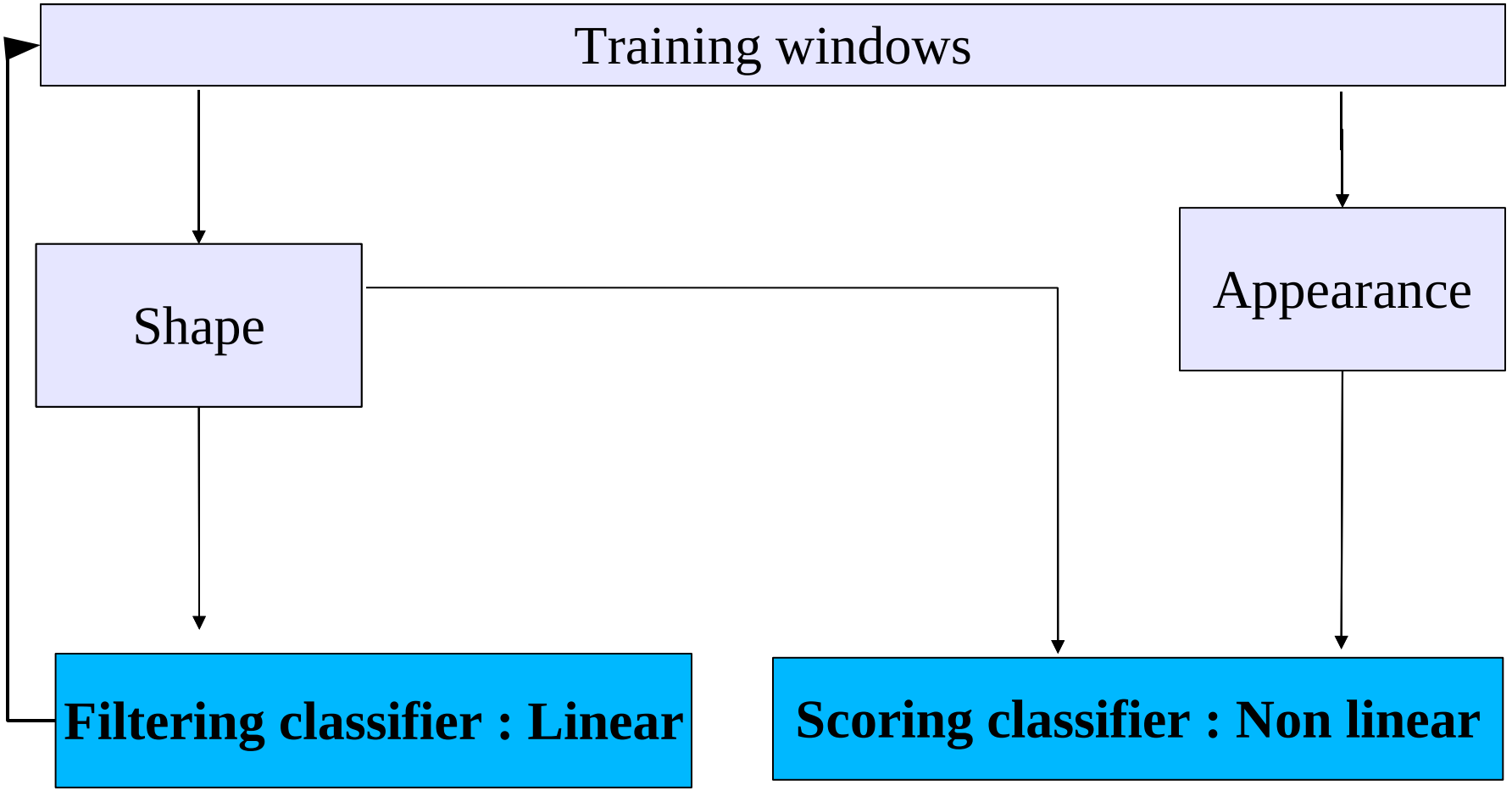
- Computed over dense patches (shift step 6 pixels, scale step 1.2)
- Discretized into 100 visual words using k-means
- Used as BOW with a spatial pyramid [Lazebnik06]



INRIA

LEAR

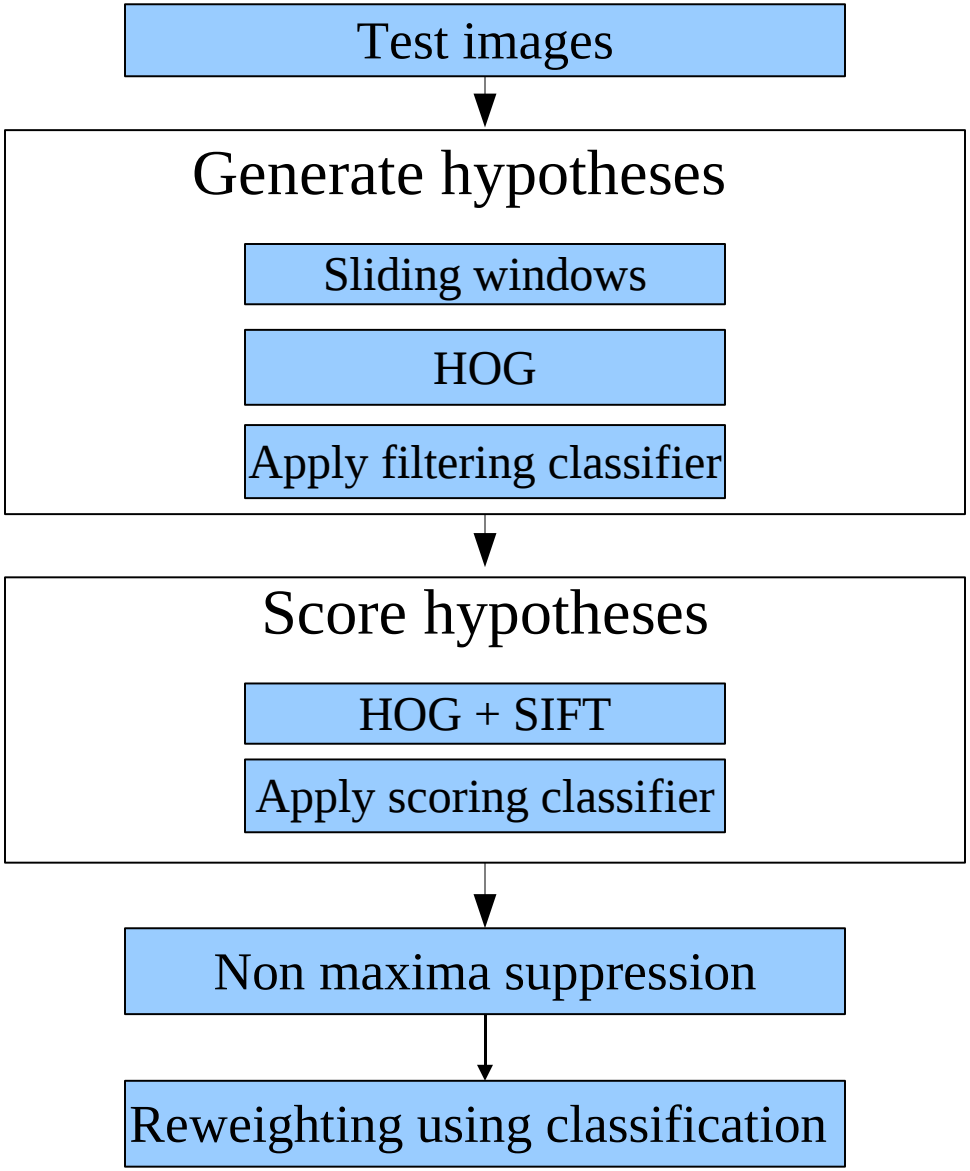
Learning procedure



Learning procedure

- Training one classifier per view (Side, Front/Rear, Unspecified)
- Linear SVM classifier
  - HOG only (combining with SIFT gives minor gain at high cost)
- Non Linear SVM classifier
  - We use non linear X^2 kernel SVM [Zhang et al 2007]
  - Training with:
    - Examples used in the linear case (positives + shifted positives + hard false positives)
    - Additional random 70K negative examples

Testing outline

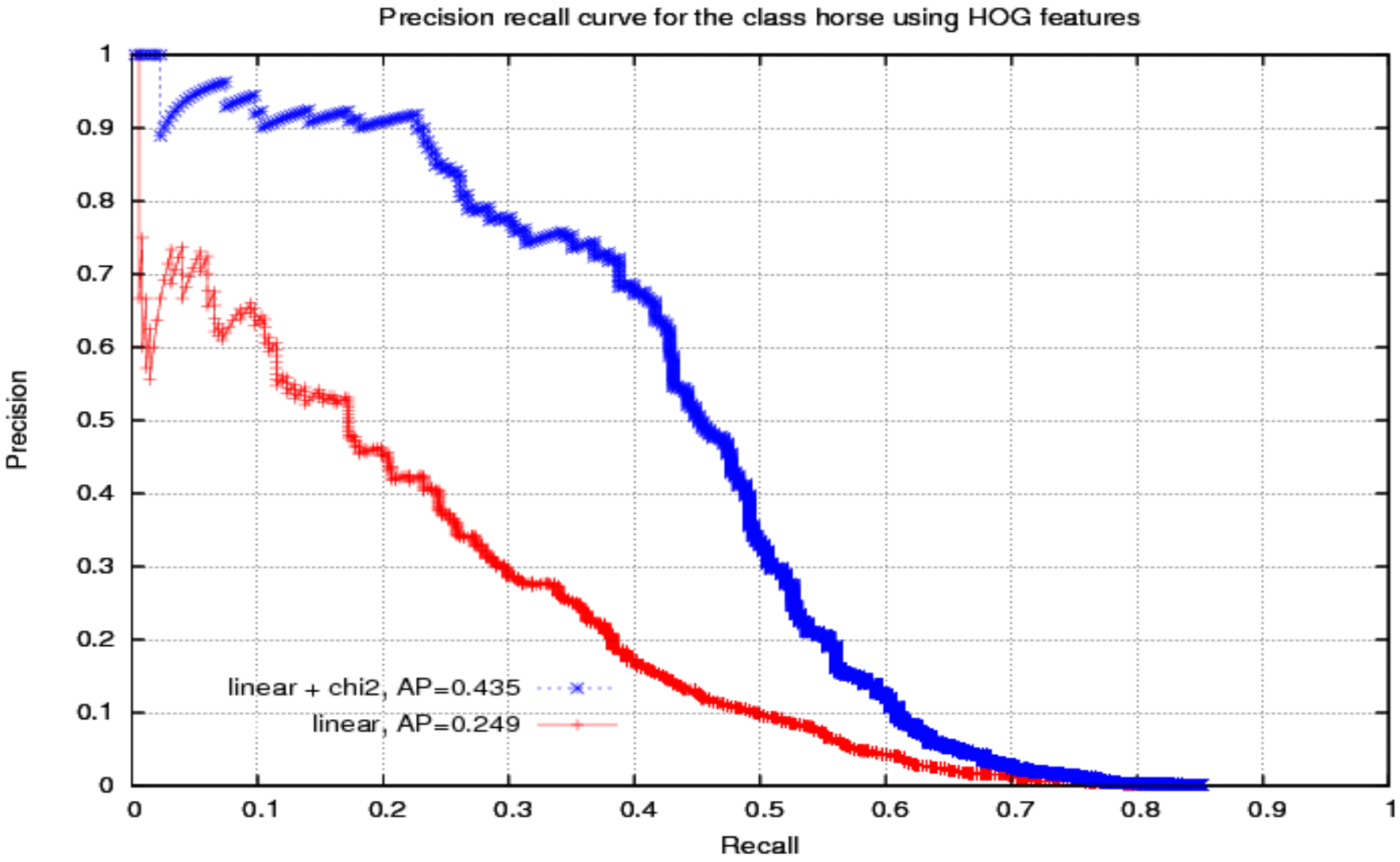


Evaluation of linear/non linear SVM

- Using HOG only to learn the non linear classifier
- Linear classifier used not only for filtering but also for scoring

	Linear	Linear + X^2
All classes	**0.139**	**0.220**
aeroplane	0.039	0.184
horse	0.249	0.435
diningtable	0.096	0.108
pottedplant	0.100	0.118

Evaluation of linear/non linear SVM



INRIA

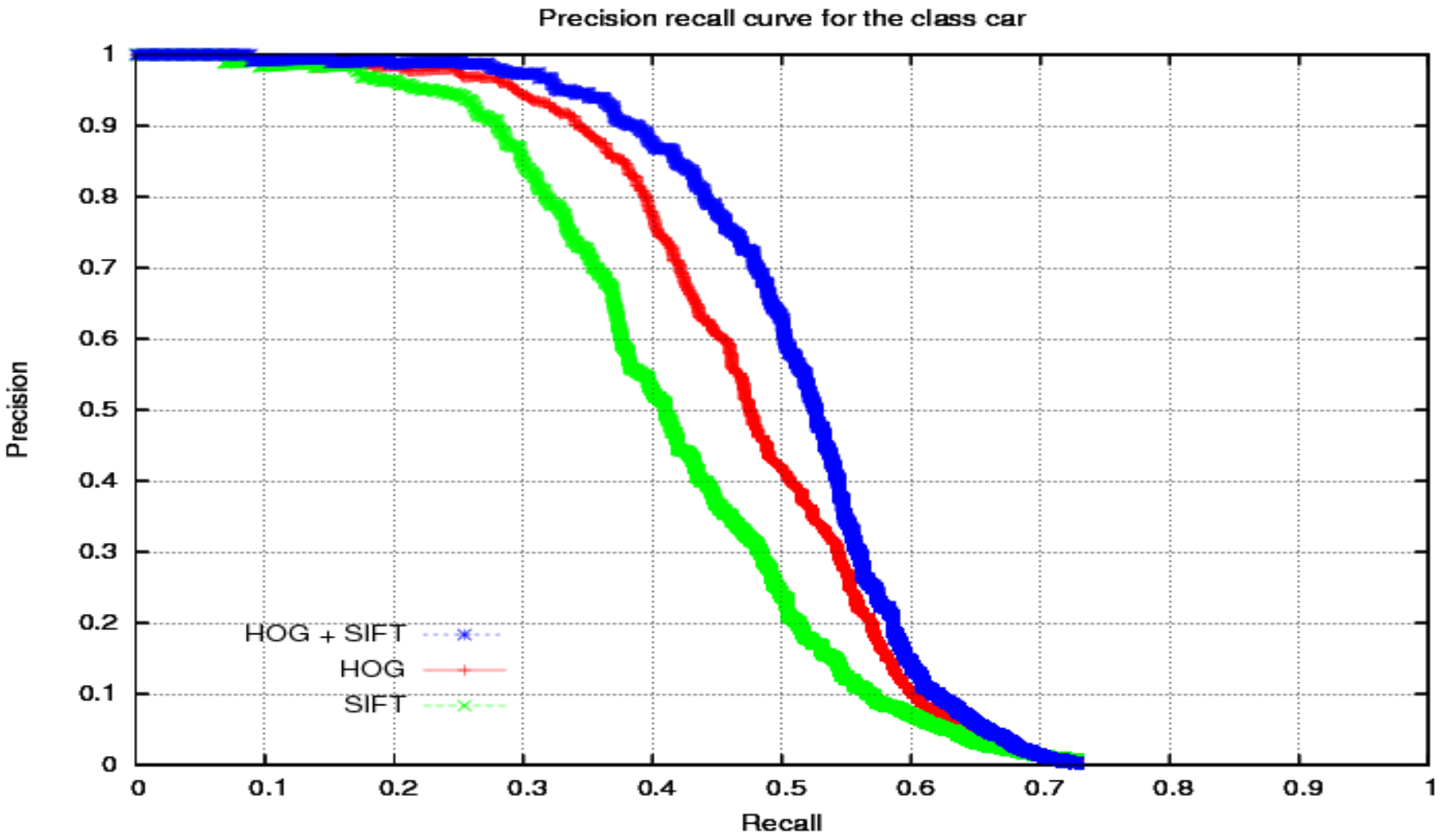
LEAR

Evaluation of descriptors

- Scoring classifiers learned with different features
- Applied on the same hypotheses

	HOG	SIFT	HOG+SIFT
All classes	**0.220**	**0.231**	**0.264**
aeroplane	0.184	0.298	0.338
car	0.475	0.425	0.511
train	0.318	0.344	0.291
bus	0.432	0.397	0.423

Evaluation of descriptors



INRIA

LEAR

Combining localization and image classification

- Provides contextual information
- Results are more reliable in image classification
- Transform scores into probabilities
- New score = P(det) * P(cls)

Influence of the use of image classification score

- We use the Lear_flat submission [Gaidon and Marszalek]

	HOG+SIFT	HOG+SIFT plusclass
All classes	**0.264**	**0.290**
cow	0.240	0.309
sheep	0.212	0.273
car	0.511	0.518
motorbike	0.417	0.427

Example of results: cars

Top true positives

First false positives

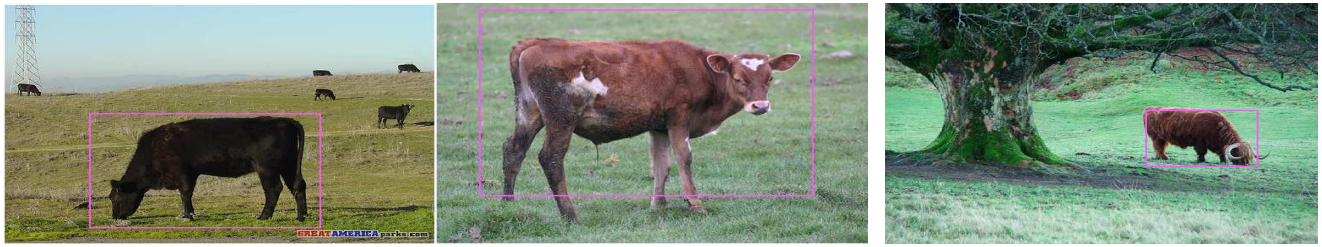
Missed

INRIA

LEAR

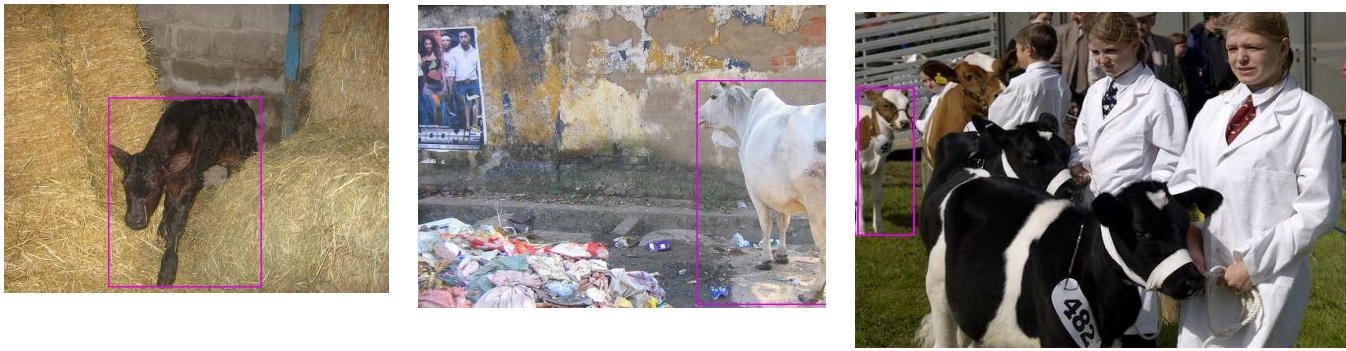
Example of results: cow

Top true positives



First false positives

Missed



INRIA

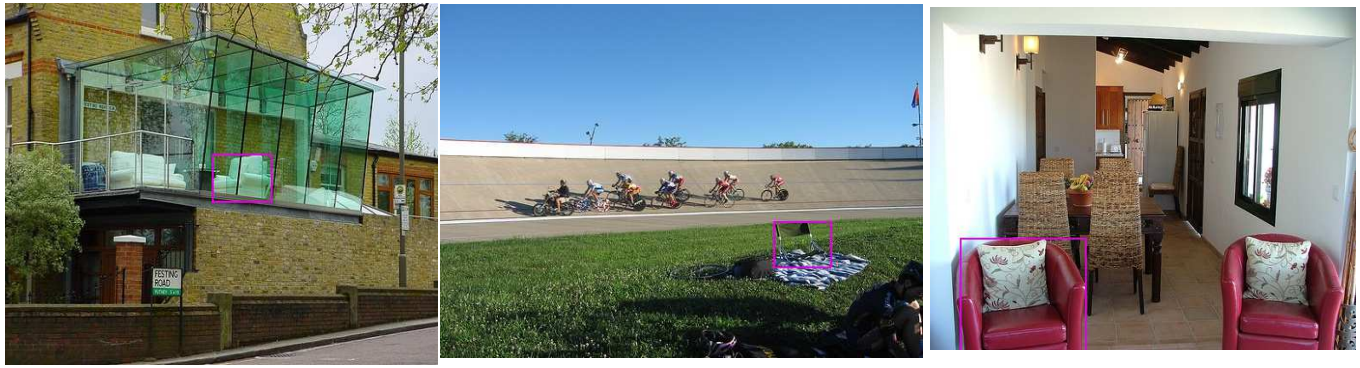
LEAR

Example of results: chair

Top true positives

First false positives

Missed



INRIA

LEAR

Summary

- Two stage classification (hypotheses prediction, hypotheses verification)
- Image representation combination
- Reweighting with classification score
- Worse performance on some articulated classes (part models more suited)
- Outperform other competitors on most of the rigid classes

Thank you